



## PATENT ABSTRACTS OF JAPAN

(11) Publication number: **11052973 A**(43) Date of publication of application: **26 . 02 . 99**

(51) Int. Cl. **G10L 3/00**  
**G06F 3/16**  
**G06F 17/21**  
**G10L 5/02**

(21) Application number: **09213566**(71) Applicant: **RICOH CO LTD**(22) Date of filing: **07 . 08 . 97**(72) Inventor: **SAKAYORI TETSUYA**(54) **DOCUMENT READING SYSTEM**

COPYRIGHT: (C)1999,JPO

## (57) Abstract:

**PROBLEM TO BE SOLVED:** To improve understandability by hierarchically capturing the contents of document from the format information and/or logical structure of the document, changing output order and/or voice attribute, etc., according thereto and changing the same to voices.

**SOLUTION:** The hierarchizing of source text is executed based mainly on tags. The part sandwiched by the tag (H1) and the tag (H2) is determined as a title and the subsequent first paragraph is determined as a content. Next, the tag of a sub-block is searched from the tags of lower layers and the section between the first tag (H2) and the next tag (H2) is determined as a sub-block. Further, from the next tag (H2) to the last are similarly processed. If there is no operation from a user at all, the title, the content and the title of the sub-block are read in this order and are shifted to the next sub-block and are similarly read. Identification is assisted by adding different effect sounds before the title, the content and the title of the sub-block, respectively and/or by changing voice kinds. The reading is continued to the lowermost layer block by repeating such.

```

<H1>
音声Webブラウザ TelMePage
</H1>
<H2>
背景
</H2>
WWWの普及が急速に進んでいます。
最近はこの技術を社内連絡用などに使
うケースも増えてビジネスの必需品と
なってきました。
<P>
しかし外出時に常に情報武装するのは
大変です。
<H2>
特徴
</H2>
TelMePageがあれば電話だけで出先から
HTML文書を確認できますので、フット
ワークとネットワークを両立できま
す。
<ul>
<li>HTMLタグを解析して内容を階層的
にブラウザ出来るので、素早いランダ
ムアクセスが可能。
<li>単語になりがちな合成音声もHTML
書式音声化機能により、わかりやすい
メッセージに変換。
</li>

```

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-52973

(43) 公開日 平成11年(1999) 2月26日

(51) Int.Cl.<sup>6</sup>  
G 1 0 L 3/00  
G 0 6 F 3/16  
17/21  
G 1 0 L 5/02

識別記号

3 3 0

F I

G 1 0 L 3/00 H  
G 0 6 F 3/16 3 3 0 H  
G 1 0 L 5/02 J  
G 0 6 F 15/20 5 6 8 Z

審査請求 未請求 請求項の数 6 O L (全 7 頁)

(21) 出願番号 特願平9-213566

(22) 出願日 平成9年(1997) 8月7日

(71) 出願人 000006747

株式会社リコー

東京都大田区中馬込1丁目3番6号

(72) 発明者 酒寄 哲也

東京都大田区中馬込1丁目3番6号 株式  
会社リコー内

(74) 代理人 弁理士 高野 明近 (外1名)

(54) 【発明の名称】 文書読み上げ方式

(57) 【要約】

【課題】 電子化文書を読み上げる際の文書の聴解性を向上させるとともに、文書読み上げ位置へのランダムアクセスを可能にする。

【解決手段】 電子化文書の表題、段落などの論理的特徴及び／又は文字の大きさなどの視覚的特徴が分かるように、前記文書の特定の情報及び／又は論理的構造によって前記文書の内容を階層化して内部データに変換し、その内部データに基づき読み上げ処理する。また、前記文書の内容を階層化した内部データを利用してユーザの指示に応じて文書の読み上げブロックに自由にアクセスできる文書読み上げシステムを提供する。

音声Webブラウザ TelMePage

背景

WWWの普及が急速に進んでいます。  
最近はこの技術を社内連絡用などに使う  
ケースも増えてビジネスの必需品となっ  
てきました。

しかし外出時に常に情報武装するのは大変  
です。

特徴

TelMePageがあれば電話だけで出先から  
HTML文書を確認できますので、フットワ  
ークとネットワークを両立できます。

- ・ HTMLタグを解析して内容を階層的にブラウズ出来るので、素早いランダムアクセスが可能。
- ・ 単語になりがちな合成音声もHTML書式音声化機能により、わかりやすいメッセージに変換。

## 【特許請求の範囲】

【請求項 1】 電子化文書の文字情報を音声合成技術によって音声化して出力する文書読み上げ方式において、前記文書の書式情報及び／又は論理的構造から当該文書の内容を階層的に捉え、これにしたがって出力文の順序及び／又は音声属性などを変更して音声化することを特徴とする文書読み上げ方式。

【請求項 2】 前記書式情報をその種類毎に異なる強さを持つ文書中の区切りとして扱い、これによって文書を階層的にブロック化することを特徴とする請求項 1 に記載された文書読み上げ方式。

【請求項 3】 文書中の第 1 段落及び段落中の第 1 文のような、前記ブロックの最初の文章単位をそれ以降の文章単位の上位の階層に位置付けることによって、前記文書の内容を階層的にブロック化することを特徴とする請求項 1 に記載された文書読み上げ方式。

【請求項 4】 前記文書の内容を上位の階層から下位の階層へ順に読み上げることを特徴とする請求項 1 に記載された文書読み上げ方式。

【請求項 5】 タイトルと内容及び階層間の識別を音声属性、付加音などを用いて行うことを特徴とする請求項 1 に記載された文書読み上げ方式。

【請求項 6】 読み出し手段がユーザの要求に応じて前記文書の上下階層間及び同一階層間で移動自在であることを特徴とする請求項 1 に記載の文書読み上げ方式。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】本発明は、SGML、HTML、などの電子構造化文書を、音声合成技術等により音声化するシステムに関し、例えば、電話でデータベースにアクセスしたり、視覚障害者が WWW (World wide Web) にアクセスする際などに利用される。

## 【0002】

【従来の技術】従来、視覚表示を前提としたテキスト(文書)の読み取り方式として次のようなものがある。

(1) テキストを音声化する際に文字装飾を音で表現したり、文書中の読み上げ位置を音像定位で表現するもの(従来技術 1 という、例えば、特開平 8-263260 号公報「テキスト読み上げ方法」)。

(2) 読み上げ位置を戻す際に文節単位毎に逆方向に、かつ、文節単位内では順方向に高速に読み上げ、利用者はこれを聞いていて聞き返したい位置を探すもの(従来技術 2 という、例えば、特開平 6-308998 号公報「音声読み上げ装置」)。

(3) 階層構造に分類された文書情報を読み上げる際、スイッチ押下時間や回数によって読み上げ位置を指定するもの(従来技術 3 という、例えば、特開平 5-281987, 281988, 281992 号公報「可搬性文書読み上げ装置」)。

(4) HTML 文書を合成音声で読み上げるもの(従来

技術 4 という、ソフトウェア製品化されている。例えば、(株)リコー製「おしゃべりさーふあー」)。

## 【0003】

【発明が解決しようとする課題】しかしながら、これら従来の文書読み上げ技術は以下で述べるようにそれぞれ聴解性及び読み上げ文書へのランダムアクセスの点で未だ十分とはいえない。この点を図面を参考にして説明すると、例えば、図 1 に示す HTML 文書は、通常ブラウザと呼ばれるソフトによって図 2 に示すように視覚的に表示される。図 2 から明らかなように、タグ情報は視覚的書式情報に変換されて文書における階層構造を明らかにしているからその視認性は向上している。ところがこれを前記従来技術 4 によって音声化すると、前記従来技術 4 では図 3 に示すようにタグを無視して日本語部分を読み上げてしまうので、ユーザにとってはだらだらと分かり難いものになってしまう。これに対し、前記従来技術 1 では音声化する際に文字修飾、文書中の読み上げ位置を表現するものであるため、ある程度視覚的特徴を音で表現できるが、文書の論理構造まで把握することは難しく、全体像の把握も音像定位という曖昧な形にとどまっている。

【0004】また、文書読み上げ方式において、音声メディアは一覧性に欠けるためランダムアクセスが難しいという問題があり、従来技術 3 では階層的ファイル構造を採用することでボタン操作によるランダムアクセスを実現している。しかしながら、従来技術 3 では専用の構造でデータを記述するため既存のテキストを対象とすることはできない。他方、従来技術 2 ではテキスト中を高速に移動することによりアクセス速度を得ているが、文書を聞き返す場合に前記文書の各文節の並びを逆の順番で聞き取りながら、文書の読み上げ位置を判断することはユーザにとって負荷が軽いとはいえない。

【0005】したがって、請求項 1 の発明の課題は、主に視覚用に作成された文書の階層構造をそのまま保存して音声化することにより、聴解性を向上させることである。請求項 2 の発明の課題は、請求項 1 の発明の課題に加え、既存の構造化された電子テキストにもそのまま適用できるようにすることである。請求項 3 の発明の課題は、請求項 1 の発明の課題に加え、構造化タグのない電子テキストにも適用できることである。請求項 4 の発明の課題は、請求項 1 の発明の聴解性を更に向上させることである。請求項 5 の発明の課題は、請求項 1 の発明の課題に加え、聴いている箇所の全体の中での位置付けを把握し易くすることである。

【0006】請求項 6 の発明の課題は、請求項 1 の発明の課題に加え、読み取り文書へのランダムアクセスを迅速に行うようにすることである。

## 【0007】

【課題を解決するための手段】本発明は、主に視覚提示用に書かれた既存の構造化された電子テキスト(文書)

を、作者の意図した文章の構造情報を含めて音声化することで、視覚表示に近い分かり易さ及びアクセスビリティの向上を目指すものである。

【0008】請求項1の発明は、電子化文書の文字情報を音声合成技術によって音声化して出力する文書読み上げ方式において、前記文書の書式情報及び／又は論理的構造から当該文書の内容を階層的に捉え、これに従って出力順序及び／又は音声属性などを変更して音声化する文書読み上げ方式である。

【0009】請求項2の発明は、請求項1の発明において、前記書式情報をその種類毎に異なる強さを持つ文書中の区切りとして扱い、これによって文書を階層的にブロック化する文書読み上げ方式である。

【0010】請求項3の発明は、請求項1の発明において、文書中の第1段落及び段落中の第1文のような、前記ブロックの最初の文章単位をそれ以降の文章単位の上位の階層に位置付けることによって、前記文書の内容を階層的にブロック化する文書読み上げ方式である。

【0011】請求項4の発明は、請求項1の発明において、前記文書の内容を上位の階層から下位の階層へ順に読み上げる文書読み上げ方式である。

【0012】請求項5の発明は、請求項1の発明において、タイトルと内容及び階層間の識別を音声属性、付加音などを用いて行う文書読み上げ方式である。

【0013】請求項6の発明は、請求項1の発明において、読みだし手段がユーザの要求に応じて前記文書の上下階層間及び同一階層間で移動自在である文書読み上げ方式である。

【0014】

【発明の実施の形態】本発明の実施態様を、図1に示すHTML文書を例に取って説明する。図4は、このHTML文書から文書の構造情報を抽出して、それを階層的内部データに変換した場合の該内部データを示したものである。図5は、その変換の際の処理フローを示したものである。以下では、まず、ソーステキストから内部データへの変換処理について述べ、続いて内部データの読み上げ処理について述べることとする。

【0015】ソーステキストの階層化は主にタグに基づいて行われる。このためにタグは例えば予め図6のように、最上層の仮想的な最上層タグから最下層の強制改行の〈BR〉まで階層が順位付けられている。この順位はタグが表わす階層の高さであり、これは論理的意味と視覚的特徴から文書の区切りとしての強さを判断して決められる。論理的意味とは表題、段落など、主に論理タグが表わす文章の意味的な構造である。視覚的特徴とは文字の大きさや野線など視覚的に感じる構造情報である。以下のブロック化では、これらの特徴をタグの持つ区切りの強さと考えて文書を階層的に分割する。なお、この実施態様及び以下の説明では主に開始タグをブロック区切りとして用い、終了タグはブロック区切りとしては使用

していないが、本発明はこれに限定されるものではなく、当然開始タグと終了タグで囲まれる範囲をブロックとして扱うことも考えられる。

【0016】以下では、図5を参考に、図1に示すソーステキストを図4に示す階層的データに変換する場合を例にとり、図5のブロック分割処理を再帰的に行うことについて説明する。まず、図1のソーステキスト全体を処理対象テキストとして、仮想的な最上層タグを現階層タグとするルートブロックのブロック化(S101～S111)を行う。タイトルと内容の抽出処理(S102)は現階層タグの種類によって異なるが、仮想的な最上層タグについては行わない。次に子ブロックCHILDのブロック化の準備として、処理対象テキストに含まれる現階層よりも下層で最も近いタグを探す(S103～S105)。ここでは第1表題タグ〈H1〉が発見される。そこで、タグ〈H1〉を対象に子ブロックをブロック化する。すなわち図1のソーステキスト全体を処理対象テキストとして、タグ〈H1〉を現階層タグとしてブロック化処理の再帰呼び出しを行う(S101)。

【0017】タグ〈H1〉についてはタグ〈H1〉とタグ〈／H1〉に挟まれる部分をタイトルとして、それに続く第1段落を内容として抽出する(S102)。その結果「音声WebブラウザTelMePage」がタイトルとして抽出され、内容は抽出されない。次に子ブロックCHILDのタグを〈H1〉より下層の候補タグから探し、第2表題タグ〈H2〉が2つ発見される(S103～S105)。そこでまず最初のタグ〈H2〉と次のタグ〈H2〉の間の部分を処理対象テキストとして、タグ〈H2〉を現階層タグとしてブロック化処理の再帰呼び出しを行い(S101～S107)、これの子ブロックとする。さらに次のタグ〈H2〉から最後までも同様に処理する(S101～S107)。

【0018】このようにして再帰的に階層ブロック化を行う(S101～S111)が、タイトルと内容を抽出する部分以外はこの処理の繰り返しとなるので、タイトル・内容抽出処理のみ以下に説明する。タグ〈H2〉ではタグ〈H1〉と同様にタイトル・内容抽出処理を行い、それに続くタグ〈P〉はタイトル無しの子ブロックとする。これは通常第1段落でそれ以降の概要などを述べることが多いことによるものである。タグ〈UL〉はここでは無視し、タグ〈L1〉部分をタイトル無しブロックとみなして処理する。このようにソーステキストから内部データへの変換が行われる。

【0019】次に、変換された内部データの読み上げ処理の一実施態様を図7を参考にして説明する。ユーザから何ら操作のない場合は、タイトル、内容、子ブロックのタイトルの順に読み上げ(S201～S210)、その子ブロックに移って同様に(タイトル、内容、子ブロックのタイトルの順に)読み上げる(S212～S214)。この時タイトル、内容、子ブロックのタイトルそ

10

20

30

40

50

れぞれの前に異なる効果音を付加するか、及び／又は声種を変えることにより識別を助けるようになっている。これを繰り返し最下層ブロックまで読み上げ（S201～S215）、その後末読ブロックに戻って読み上げを続ける（S202～S215）。これによってこのテキストが全体としては背景と特徴からなることが読み上げの冒頭で分かり聴解性が向上する。

【0020】さらに読み上げ途中でユーザからの割り込みによるロケーション指定を受け付けることができ、これによりランダムアクセスが可能となる。即ち、「もう一度」を指示することで（S301）現在読み上げているブロックの先頭へ返って聴き返すことができる（S203～S214）。また、例えばタグ〈H1〉ブロックの子ブロックタイトル「特徴」を読み上げている時に、「下層へ」を指示すると（S302）、読み上げロケーションを直接「特徴」の中身へ飛ばす（S212）ことが出来る。タグ〈H2〉「特徴」の内容の読み上げ中に「次へ」を指示すると（S303）、親ブロックの次の子ブロック、すなわちタグ〈H2〉ブロック「背景」へ飛ぶことが出来る。箇条書き部分の（U）ブロックについても同様に読み上げ中に次の項目に飛ぶことができる。「上層へ」を指示すると（S304）親ブロックの先頭へ返って読み上げることができる。

【0021】以上の読み上げ動作は図7の流れ図で示される処理（S201～215）を再帰的に用いて実現することができる。図中、blockTITLEはブロックblockのタイトル、blockCONTENTはブロックblockの内容、blockNCHILDはブロックblockの子ブロックの数、blockCHILD[i]はブロックblockのi番目の子ブロックをそれぞれ表している。なお、テキストの音声化処理については既存のテキスト音声合成技術が使えるのでここでは説明を省略する。

【0022】

【発明の効果】

\*

【図3】

音声Webブラウザ TelMePage。背景。WWWの普及が急速に進んでいます。最近はこの技術を社内連絡用などに使うケースも増えてビジネスの必需品となりました。しかし外出時に常に情報武装するのは大変です。特徴。TelMePageがあれば電話だけで出先からHTML文書を確認できますので、フットワークとネットワークを両立できます。HTMLタグを解析して内容を階層的にブラウザ出来るので、素早いランダムアクセスが可能。単調になりがちな合成音声もHTML書式音声化機能により、わかりやすいメッセージに変換。

\* 請求項1に対応する効果：文書の情報を、その階層構造を保存したまま話し言葉としての表現に写像して出力することができるため、聴解性を向上することができる。請求項2に対応する効果：請求項1に対応する効果に加えて、既存の構造化された電子テキストにもそのまま適用することができる。

請求項3に対応する効果：請求項1に対応する効果に加えて、既存の構造化タグのない電子テキストにも適用することができる。

10 請求項4に対応する効果：請求項1に対応する効果に加えて、下位概念を聴解するための前提知識である上位概念を常に事前に取得することになるため全体的に聴解性が向上する。

請求項5に対応する効果：請求項1に対応する効果に加えて、現在聞いている個所の文書全体の中での位置づけが把握しやすく、聴解性が向上する。

20 請求項6に対応する効果：請求項1に対応する効果に加えて、音声メディアの欠点であるランダムアクセスの難しさを補い、ユーザ要求に対する反応速度を向上することができる。

【図面の簡単な説明】

【図1】 読み上げ文書の例を示す図である。

【図2】 前記文書のブラウザによる表示例を示す図である。

【図3】 前記文書を従来の方式で音声化した場合の読み上げ表示例を示す図である。

【図4】 前記文書の階層的内部データを示す図である。

30 【図5】 前記文書の構造を階層的内部データに変換する処理フローを示す図である。

【図6】 前記文書に付与されるタブの一例を示す図である。

【図7】 階層的内部データによる文書読み上げ処理フローを示す図である。

【図6】

上	仮想的最上層タグ
↑	〈H1〉 表題1
	〈HR〉 野線
	〈H2〉 表題1
	〈H3〉 表題1
階層	〈P〉 段落
↓	〈L1〉 箇条書き
	〈TR〉 表の行
下	〈BR〉 強制改行

【図1】

```

<H1>
音声Webブラウザ TelMePage
</H1>
<H2>
背景
</H2>
WWWの普及が急速に進んでいます。
最近はこの技術を社内連絡用などに使う
ケースも増えてビジネスの必需品と
なってきました。
<P>
しかし外出時に常に情報武装するのは
大変です。
<H2>
特徴
</H2>
TelMePageがあれば電話だけで出先から
HTML文書を確認できますので、フット
ワークとネットワークを両立できま
す。
<ul>
<li>HTMLタグを解析して内容を階層的
にブラウズ出来るので、素早いランダ
ムアクセスが可能。
<li>単調になりがちな合成音声もHTML
書式音声化機能により、わかりやすい
メッセージに変換。
</ul>

```

【図2】

## 音声Webブラウザ TelMePage

### 背景

WWWの普及が急速に進んでいます。  
最近はこの技術を社内連絡用などに使う  
ケースも増えてビジネスの必需品となっ  
てきました。

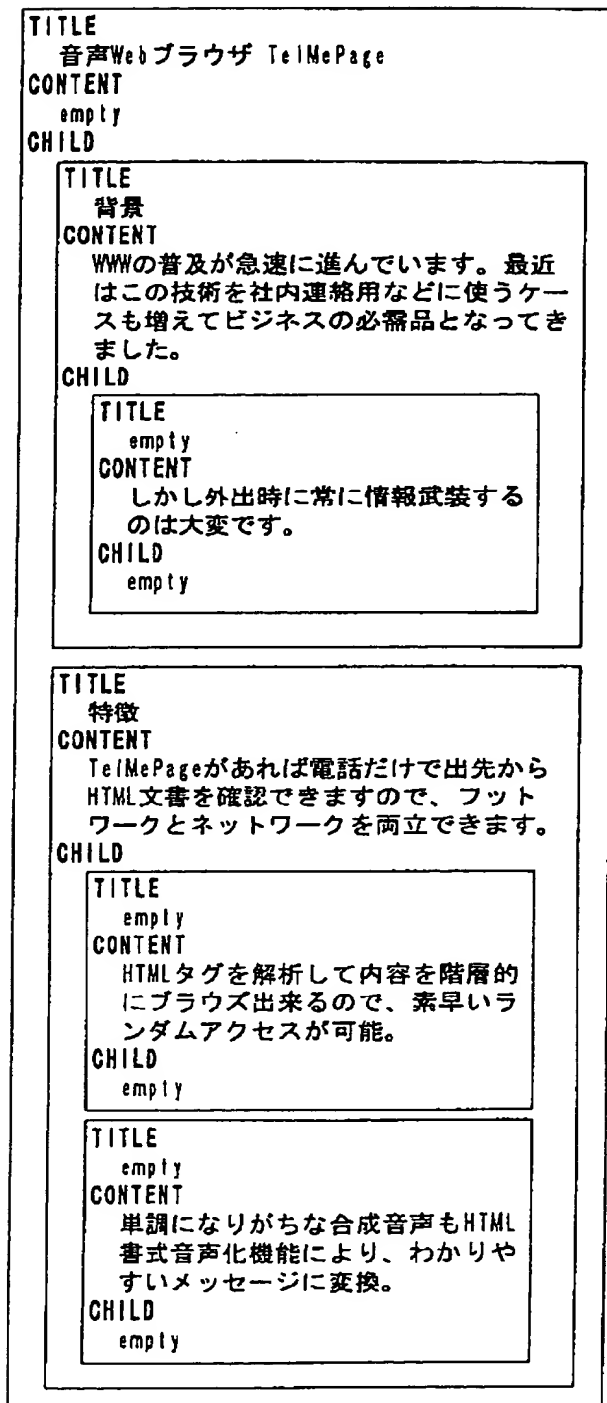
しかし外出時に常に情報武装するのは大変  
です。

### 特徴

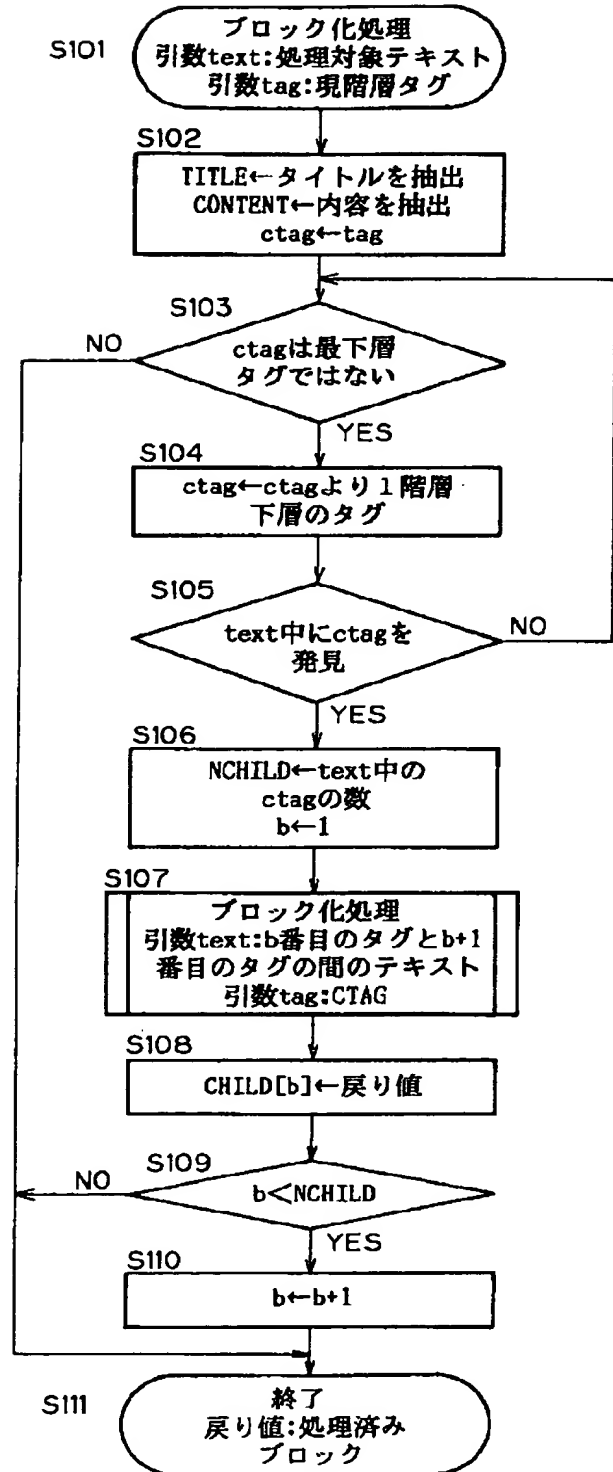
TelMePageがあれば電話だけで出先から  
HTML文書を確認できますので、フットワ  
ークとネットワークを両立できます。

- ・HTMLタグを解析して内容を階層的にブラウズ出来るので、素早いランダムアクセスが可能。
- ・単調になりがちな合成音声もHTML書式音声化機能により、わかりやすいメッセージに変換。

【図4】



【図5】



【図7】

